New Researches in the Smart City
Vol 3 (Issue 1), 29-49
nrsc.apadana.ac.ir

Apadana Institute of Higher Education

Extracting Appropriate Features for Behavior Analysis and Predicting Customer Attrition in Banks

Semiramis Namini *

Master of Science, Faculty of Engineering,
Department of Computer Engineering - Software,
Khajeh Nasir Toosi University of Technology,
Tehran, Iran.

Received: 02/September/2024
Accepted: 15/November/2024
eISSN: 3060-6144
ISSN: 2980-8936

Abstract

Customer loyalty has always been one of the main concerns and challenges for domestic banks, as their lifecycle depends on customer loyalty. Various studies have been conducted on customer loyalty in banks and financial and commercial institutions. This research evaluated the use of data mining and five machine learning algorithms—Support Vector Machine, Naive Bayes, MLP, Decision Tree, and Logistic Regression—for detecting customer loyalty in a domestic bank. Three performance metrics—accuracy, precision, and recall—were used to compare the results. The data used in this research were collected from a domestic bank and consisted of 15 criteria for evaluating customer loyalty to the organization. The results showed that, among the five algorithms, the best performance was achieved by the Decision Tree, while the Logistic Regression algorithm showed the worst performance. Finally, to improve the results, a combination of a neural network and the Teacher-Student optimization algorithm was used to achieve better detection of loyal customers.

Keywords: customer loyalty, data mining, decision tree, bank

* Corresponding Author: S.nameny62@gmail.com

فصلنامه پژوهش‌های نوین در شهر هوشمند
دوره ۳، شماره ۱، پاییز ۱۴۰۳، ۲۹-۴۹
nrsc.apadana.ac.ir

مؤسسه آموزش عالی آپادانا
غیر دولتی - غیر انتفاعی

تاریخ دریافت: ۱۴۰۳/۰۶/۱۲ تاریخ پذیرش: ۱۴۰۳/۰۸/۲۵ eISSN: 3060-6144 ISSN: 2980-8936

استخراج ویژگی‌های مناسب جهت تحلیل رفتار و پیش‌بینی دلایل ریزش مشتریان بانک‌ها

سمیرامیس نامنی *
کارشناسی ارشد، دانشکده فنی و مهندسی، گروه مهندسی کامپیوتر-نرم افزار، دانشگاه خواجه نصیر، تهران، ایران.

چکیده

بررسی وفاداری مشتریان، همواره از دغدغه‌ها و چالش‌های بانک‌های داخلی به شمار می‌رود چراکه چرخه حیات آن‌ها به وفاداری مشتریان بستگی دارد. پژوهش‌های متفاوتی در زمینه بررسی وفاداری مشتریان بانک‌ها و مؤسسات مالی و تجاری انجام شده است. در این پژوهش نیز روش داده‌کاوی و پنج الگوریتم ماشین بردار پشتیبان، نایوبیز، MLP، درخت تصمیم و رگرسیون لجستیک در تشخیص وفاداری مشتریان بانک داخلی مورد ارزیابی قرار گرفت. همچنین از سه معیار دقت، صحت و فراخوانی برای مقایسه نتایج به‌دست‌آمده استفاده گردید. داده‌های مورد استفاده در این پژوهش از یک بانک داخلی جمع‌آوری شد که شامل ۱۵ معیار برای ارزیابی وفاداری مشتریان به این سازمان بود. نتایج نشان داد که در بین پنج الگوریتم ارائه‌شده، بهترین عملکرد مربوط به درخت تصمیم و بدترین عملکرد نیز مربوط به الگوریتم رگرسیون لجستیک است. درنهایت برای بهبود عملکرد از تلفیق شبکه عصبی با الگوریتم بهینه‌سازی معلم دانش‌آموز استفاده شد تا بتوان از نتایج بهتری برای تشخیص مشتریان وفادار استفاده نمود.

کلیدواژه‌ها: وفاداری مشتریان، داده‌کاوی، درخت تصمیم، بانک

* نویسندهٔ مسئول: S.nameny62@gmail.com

۱- مقدمه

مدیریت ارتباط با مشتری[۱]، یک ابزار مدیریتی مدرن بوده که از تجزیه‌وتحلیل داده‌ها و داده‌کاوی[۲] با هدف جذب مشتریان و رساندن مشتری به ارزش‌های لازم جهت حفظ و نگهداری مشتری استفاده می‌کند. دیدگاه مشتری به شما کمک می‌کند تا شکاف‌ها و فرصت‌های موجود را ببینید و استراتژی و فرآیندهای جدید و مناسبی را برای سازمان خود اتخاذ کنید. مدل مدیریت دانش مشتری[۳]، اخیراً از طریق همگرایی هر دو رویکرد مبتنی بر فناوری و داده‌محور در CRM و رویکرد مردم‌مدار در مدیریت دانش و تعامل آن‌ها مورد توجه قرار گرفته است. دانش مشتری، دستیابی به اطلاعات و بینشی است که برای برقراری ارتباط قوی‌تر با مشتریان به آن نیاز است. انتظار از این تلاش، به دست آوردن دانشی جامع‌تر برای مشتریان، دانش در مورد مشتریان و دانش از مشتریان است.

داده‌کاوی به‌عنوان فرایندی تعریف می‌شود که از روش‌های ریاضی[۴]، آماری[۵]، هوش مصنوعی[۶] و یادگیری ماشینی[۷] به‌منظور استخراج و شناسایی اطلاعات مفید و متعاقباً کسب دانش از پایگاه‌های داده استفاده می‌کند. از الگوریتم‌های داده‌کاوی در طیف وسیعی از زمینه‌های تحقیقاتی مانند مراقبت‌های بهداشتی و پزشکی (Zhou et al., 2015)، تجزیه‌وتحلیل احساسات (Tao et al., 2016; Zhou et al., 2017)، آموزش (Abdar et al., 2018) و غیره استفاده شده است. هدف از کاربرد داده‌کاوی در صنعت بانک، استفاده از داده‌های موجود برای حفظ بهترین مشتریان و شناسایی فرصت‌ها جهت فروش خدمات اضافی به آن‌هاست.

ابزارهای فناوری اطلاعات و داده‌های بسیار زیاد مشتریان بانک‌ها می‌تواند در بهبود روابط مدیریتی بین بانک‌ها و مشتریان بسیار مفید باشد. داده‌کاوی، ابزاری است که در بانک‌ها و سازمان‌های مالی برای تصمیم‌گیری و پیش‌بینی مورد استفاده قرار می‌گیرد. طبقه‌بندی[۸]، یکی از رایج‌ترین مدل‌های یادگیری در داده‌کاوی است که رفتارهای آینده مشتریان را پیش‌بینی می‌کند. پیش‌بینی با طبقه‌بندی پایگاه داده به چندین کلاس از پیش تعیین‌شده، بر اساس معیارهای خاصی انجام می‌گردد. شبکه‌های عصبی[۹]، درختان تصمیم‌گیری[۱۰]، بیز ساده[۱۱]، رگرسیون لجستیک[۱۲]، قانون وابستگی[۱۳] و ماشین بردار پشتیبانی[۱۴]، ابزارهای متداول مورد استفاده برای طبقه‌بندی به شمار می‌روند (Bahari & Elayidom, 2015).

در این تحقیق، یک چارچوب جدید مدیریت دانش مشتری[۱۵] مبتنی بر داده‌کاوی پیشنهاد شده است. در ابتدا، با استفاده از تکنیک‌های داده‌کاوی استاندارد و سپس با تلفیق شبکه عصبی با الگوریتم بهینه‌سازی معلم دانش‌آموز[۱۶]، به تشخیص پیش‌بینی وفاداری مشتریان به بانک می‌پردازیم. علاوه بر این، برای کاربردهای CRM در حوزه بانکی، تقسیم‌بندی مشتریان، جستجو و دستیابی، امنیت، سودآوری، تجزیه‌وتحلیل ریسک و نقاط قوت و ضعف برای نشان دادن عملکرد مدل‌های طبقه‌بندی در نظر گرفته شده است.

---

1. Customer relationship management (CRM)
2. data mining
3. Customer knowledge management (CKM)
4. mathematical
5. statistical
6. artificial intelligence
7. machine learning
8. classification
9. neural networks
10. decision trees
11. naive bayes
12. logistic regression
13. association rule
14. support vector machine
15. CKM
16. Teaching learning base optimization (TLBO)

۲- پیشینه پژوهش

۲-۱- مدیریت ارتباط با مشتری

در پژوهش مانیانگا و همکاران[1] (۲۰۲۲) بیان شده است که تکنیک‌های E-CRM از نظر مدیریت ارتباط با مشتری مؤثر است. علاوه بر این، یافته‌ها نشان داد که E-CRM تأثیر مطلوبی بر حفظ مشتری دارد و در بهبود عملکرد و ارزش کلی شرکت نیز حیاتی و مؤثر است. به بانک‌ها توصیه شد که منابعی را برای استفاده از برنامه‌های کاربردی مدرن E-CRM، معیارسازی و اجرای بهترین تکنیک‌های E-CRM و ایجاد فرهنگ‌های E-CRM اختصاص دهند. همچنین، دش و نایاک[2] (۲۰۲۲) با بررسی CRM در بانک‌های هندی به این نتیجه رسیدند که بانک‌ها می‌توانند هدف کسب‌وکار خود را تحقق بخشند و مشتریان ارزشمندی را به دست آورند و حفظ کنند تا آن‌ها را در کل فرآیند کسب‌وکار مرتبط کنند. باید یک تلاش بسیار یکپارچه برای نوآوری، کشف، ایجاد و برآوردن نیازهای مشتری به شکلی متفاوت انجام شود. بااین‌حال، ذکر این نکته شگفت‌آور است که بیشتر فعالیت‌های بخش بانکی و مالی به‌جای حفظ مشتری، بر جذب مشتری متمرکز است. کومار و موخا[3] (۲۰۲۲) به بررسی تأثیر E-CRM بر وفاداری مشتری با تأثیر میانجی رضایت مشتری در صنعت بانکداری پرداختند. نتایج این پژوهش نشان داد که E-CRM و رضایت مشتری، تأثیر مثبت معناداری بر وفاداری مشتری دارند و رضایت مشتری، تا حدی رابطه بین E-CRM و وفاداری مشتری را واسطه می‌کند. در پژوهش کا و جانسون[4] (۲۰۱۷) بیان شده است که مدیریت ارتباط با مشتری الکترونیکی (e-CRM) به آخرین پارادایم در دنیای مدیریت ارتباط با مشتری تبدیل شده است. با ورود تجارت به وب، e-CRM بیشتر و بیشتر ضروری می‌شود. امروزه، رضایت مشتری هدف اصلی در صنعت بانکداری است. با پیشرفت در فناوری اطلاعات و ارتباطات در هزاره جدید، اکنون فعالیت‌های بانکی تنها به استقراض و وام محدود نمی‌شود بلکه خدمات فراوانی را ارائه می‌دهد. در محیط بانکداری در حال تغییر سریع در سراسر جهان، بانک‌های هند نه‌تنها باید قوانین جدید را بیاموزند بلکه باید مهارت‌ها و ابزارهای بانکداری را نیز ارتقاء دهند.

۲-۲- داده‌کاوی برای CRM

در پژوهش دس و نایاک[5] (۲۰۲۲) بیان شده است که مشتریان در جامعه هوشیارتر، باهوش‌تر و پویاتر هستند. آن‌ها ترجیحات و عادات خود را با توجه به نیازهای خود تغییر می‌دهند. شناخت نیازهای مشتریان، بخش مهمی از بازاریابی است که در آن یک شرکت باید مشتریان وفادار را در این ناهمگونی کشف کند. در پژوهش رضایی و همکاران (۲۰۲۲) بیان شده است که بازاریابی چند کاناله باعث می‌شود مشتری در کانال‌های مختلف، هویت منحصربه‌فردی نداشته باشد. این موضوع، هم‌افزایی کانال‌ها را در تقویت نگرش مثبت مشتریان تحت‌الشعاع قرار می‌دهد. بااین‌حال، یک استراتژی بازاریابی همه‌کانالی می‌تواند به‌درستی کار کند. در پژوهش لطیف و وینیتا[6] (۲۰۲۱b) بیان شده است که سیستم‌های مدیریت ارتباط با مشتری برای توانمندسازی سازمان‌ها به‌منظور به دست آوردن مشتریان جدید، ایجاد ارتباط مستمر با آن‌ها و افزایش حفظ مشتری برای سودآوری بیشتر استفاده می‌شود. وفاداری مشتری به‌عنوان ریزش مشتری نیز شناخته می‌شود. در پژوهش مذکور با استفاده از یک مدل ترکیبی با ترکیب مدل پرسپترون چندلایه (MLP) و مدل شبکه عصبی کانولوشنال[7] با دقت ۸۷ درصد، میزان ریزش مشتریان پیش‌بینی گردید. هدف پژوهش

1. Manyanga et al.
2. Dash & Nayak
3. Kumar & Mokha
4. KA & Johnson
5. Das & Nayak
6. Latheef & Vineetha
7. CNN

لطیف و وینیتا (۲۰۲۱a)، شناسایی و تجسم عواملی است که در ریزش مشتری نقش دارند. بر این اساس، یک مدل پیش‌بینی با استفاده از الگوریتم یادگیری گروهی ایجاد می‌کند. این پژوهش با استفاده از زبان برنامه‌نویسی پایتون پیاده‌سازی گردید و نتایج توانست با دقت ۸۶ درصد میزان ریزش مشتریان را پیش‌بینی نماید. در کل می‌توان بیان نمود مفهوم وفاداری برای یک کسب‌وکار در زمینه بازاریابی از اهمیت فوق‌العاده‌ای برخوردار است تا ماهیت تمایل مشتریان برای ادامه حمایت طولانی‌مدت از یک شرکت را توصیف کند. وفاداری مشتری از دیدگاه‌های مختلف و در زمینه‌های مختلف مانند وفاداری به برند، وفاداری فروشنده، وفاداری به خدمات و وفاداری فروشگاه، به دلیل میزان تأثیر و علاقه آن در بین متخصصان مورد توجه قرار گرفته است. وفاداری مشتری به یک ابزار استراتژیک در یک زمینه تجاری بسیار پویا و رقابتی تبدیل شده است. با توجه به بزرگی تأثیر آن، وفاداری به‌عنوان یک شعار جدید برای رشد پایدار و سودآوری یک تجارت تلقی می‌شود.

۳- داده‌کاوی

اصطلاح داده‌کاوی برگرفته از استخراج طلا[1] از صخره‌های سنگی است. در رابطه با استخراج و اکتشاف طلا از واژه Rock mining استفاده نشده است؛ بنابراین، شاید نام مناسب‌تر برای Data mining نیز واژه Knowledge mining from data بود اما ازآنجاکه این واژه طولانی است و ممکن است اصطلاح کوتاه‌تر Knowledge mining نیز نتواند مؤکد بر استخراج دانش از میزان وسیعی از داده‌ها باشد، از واژه Data mining استفاده می‌شود (Mining, 2006). داده‌کاوی مفهومی است که با پایگاه‌های داده‌های بزرگی مانند Data repositoryها و Data warehouseها عجین شده است (Wager et al., 2005) و به دنبال استخراج اطلاعات مفید و ناشناخته از داده‌های خام است (Tan, 2008). اگرچه داده‌کاوی همچون دیگر مفاهیم فناوری اطلاعات، معانی متعددی را برای افراد مختلف تداعی می‌کند اما اگر به‌طور دقیق استفاده شود، به ابزار تحلیلی پیچیده‌ای اشاره دارد که به‌طور خودکار الگوهای مفیدی را در میان داده‌های یک مخزن داده کشف می‌کند. درواقع، داده‌کاوی شکل پیشرفته پشتیبانی از تصمیم است و برخلاف ابزارهای پرس‌وجوی غیرفعال[2]، بدون الزام به طرح سؤال از طرف کاربر به تولید الگو، روندها و قواعد برنامه‌ریزی‌شده می‌پردازد (Wager et al., 2005). به عبارت دیگر، قدرت داده‌کاوی در این است که می‌تواند الگوهایی را که در جست‌وجوی کاربر مورد توجه قرار نگرفته است، افشا کند و پاسخ‌هایی را برای سؤالاتی که هرگز درخواست نشده بود، تولید نماید (LaTour & Eichenwald, 2010). بدین ترتیب، هدف داده‌کاوی، استخراج اطلاعات باارزش از داده‌ها با هدف غایی کشف دانش است (Tan, 2008). لذا، برخی از افراد، اصطلاح کشف دانش از داده‌ها[3] را به‌جای داده‌کاوی به کار می‌برند و برعکس، عده زیادی داده‌کاوی را هسته فرایند کشف دانش از داده‌ها و گامی اساسی در آن دانسته (Mining, 2006) و آن را یکی از مهم‌ترین زیرشاخه‌های مدیریت دانش معرفی می‌کنند (Chen et al., 2006). با توجه به تعریف داده‌کاوی، نظر دوم صحیح است و داده‌کاوی بخشی از کشف دانش و شامل مراحل زیر است:

- پاک‌سازی داده‌ها[4]: در این مرحله، داده‌های غیر همسان و زائد از میان داده‌ها خارج می‌شود.
- یکپارچگی داده‌ها[5]: در این مرحله، منابع متعدد داده‌ها با هم ترکیب می‌شوند.
- گزینش داده‌ها[6]: در این مرحله، داده‌های مربوط به هدف، از پایگاه داده‌ها بازیابی می‌شود.

1. gold mining
2. passive
3. knowledge discovery from data
4. data cleaning
5. data integration
6. data selection

- تغییر شکل داده‌ها[1]: در این مرحله، داده‌ها تغییر شکل یافته و به‌واسطه عملیات مختلفی مثل خلاصه‌برداری، به اشکال مناسب برای استخراج تبدیل می‌شوند.
- داده‌کاوی: فرایندی ضروری است که برای استخراج الگوهای داده‌ها از شیوه‌های منطقی و خردمند استفاده می‌کند.
- ارزشیابی الگو[2]
- ارائه دانش[3]

هر یک از این مراحل، پیش‌نیاز مرحله بعد بوده و فرایندی تکرارپذیر است. بدین ترتیب، کل فرایند کشف دانش از داده‌ها، از مجموعه وسیعی از داده‌ها شروع شده و به محصول نهایی دانش و ایجاد مدل ختم می‌شود (Godin, 2002). داده‌کاوی با سرنام SEMMA مشخص می‌شود. این سرنام متشکل از نمونه[4]، بیانگر یک نمونه آماری از داده‌ها، اکتشاف[5]، بیانگر استفاده از تکنیک‌های اکتشافی و آماری، اصلاح[6] یا دست‌کاری[7]، بیانگر گزینش و تغییر شکل مهم‌ترین متغیرهای قابل پیش‌بینی، مدل[8] به معنای مدلی از متغیرها برای پیش‌بینی پیامدها یا الگوها و ارزیابی[9] به معنای تأیید صحت مدل، است. SEMMA، خود یک چرخه است که در صورت نیاز می‌تواند به‌طور متوالی انجام شود (Obenshain, 2004).

۴- روش

روش پیشنهادی این پژوهش، بهبود عملکرد شبکه عصبی به‌عنوان یک روش ارائه‌شده برای داده‌کاوی است. در این بخش به بیان روش اجرای این پژوهش پرداخته شده است که در بخش نخست به ارائه ساختار شبکه عصبی به‌عنوان غلاف مرکزی تحلیل و در بخش بعد به بیان الگوریتم بهینه‌سازی و نیز موارد مرتبط به آن به‌عنوان الگوریتمی در دل شبکه عصبی برای بهینه‌سازی اوزان پرداخته شده است و در انتها نیز فلوچارت لحظه ورود دیتاست به تحلیل تا حصول نتایج ارائه شده است.

۴-۱- ساختار شبکه عصبی

در این بخش به بیان ساختار شبکه عصبی مورد استفاده در این پژوهش پرداخته شده و موارد زیر به ترتیب بیان خواهد شد.

- نوع شبکه
- ورودی‌ها
- خروجی‌ها
- تابع-آموزش
- تابع عملکرد
- تعداد لایه‌های پنهان

1. data transformation
2. pattern evaluation
3. knowledge presentation
4. sample
5. explore
6. modify
7. manipulation
8. model
9. assess

- تابع فعال‌ساز
- تعداد تکرار
- نوع شبکه

در این تحلیل از یک شبکه از نوع شبکه پیش‌خور[1] استفاده شده است که در این روش تحلیل روبه‌جلو صورت گرفته و نورون‌ها به سمت عقب برای دریافت داده بازنمی‌گردند و همواره به جلو حرکت دارند؛ بنابراین، بازخورد یا فیدبک وجود ندارد، به این معنا که خروجی هر لایه تنها بر لایه بعد اثر می‌گذارد و تغییری در لایه خودش ایجاد نمی‌کند.

- ورودی

در الگوریتم ارائه‌شده از یک مجموعه ورودی دریافت‌شده از دیتاست مورد استفاده برای این پژوهش استفاده شده است که در آن چندین ویژگی آموزش به شبکه معرفی خواهد شد.

- خروجی

خروجی‌ها یا اهداف شبکه نیز پاسخ نهایی ارائه‌شده در نمونه‌های واقعی موجود در دیتاست هستند. تعداد این اهداف که همان دسته‌بندی‌ها را مشخص می‌نماید، دو عدد مختلف است.

- تابع آموزش

در این الگوریتم از میان روش‌های مختلف آموزش از روش لونبرگ ماکوارت استفاده شده است. این الگوریتم روشی است برای یافتن کمینه یک تابع غیرخطی چندمتغیره که به‌عنوان یک روش استاندارد برای حل مسئله کمینه مربعات برای توابع غیرخطی درآمده است. برای این منظور، نرم‌افزار متلب بسته آماده‌ای داشته که شبکه را می‌توان با آن آموزش داد. این کار با فراخوانی دستور trainlm در مکان مناسب انجام می‌گردد.

- تابع عملکرد

بعد از مشخص شدن اوزان هر یک از متغیرها و ویژگی‌ها، باید بررسی شود این اوزان درست انتخاب شده‌اند یا خیر. برای این منظور از توابع عملکرد استفاده شده است. در این پژوهش از مربع خطا به‌عنوان تابع عملکرد کمک گرفته شده است. خطا در این پژوهش، میزان اختلاف خروجی به‌دست‌آمده از مقادیر واقعی موجود در دیتاست بوده و طبیعتاً بهترین مقدار برابر صفر است. این رابطه در ادامه بیان شده است.

$$MSE = \frac{\sum\left(T - Network(\acute{T})\right)^2}{N} \quad (۱)$$

در رابطه بیان‌شده، T بیانگر مقادیر هدف موجود در دیتاست، I بیانگر ورودی‌های حال حاضر الگوریتم در دور به‌دست‌آمده بوده که به‌عنوان متغیر به شبکه معرفی شده و میزان به‌دست‌آمده از مقدار T کاسته می‌شود.

- تعداد نورون‌ها

تعداد نورون‌ها همان تعداد جواب‌های محتمل است و هرچه بیشتر باشد، برای دقت شبکه بهتر بوده اما از سرعت تحلیل کاسته خواهد شد. در این پژوهش، تعداد نورون‌ها برابر تعداد متغیرهای ورودی در نظر گرفته شده است.

- تعداد تکرار

تعداد تکرار در این الگوریتم برابر عددی است که الگوریتم بعد از رسیدن به همگرایی از آن خارج می‌شود. در این زمینه، الگوریتم آزادانه عمل نموده و تا رسیدن به همگرایی کامل از چرخه بهینه‌سازی خارج نخواهد شد.

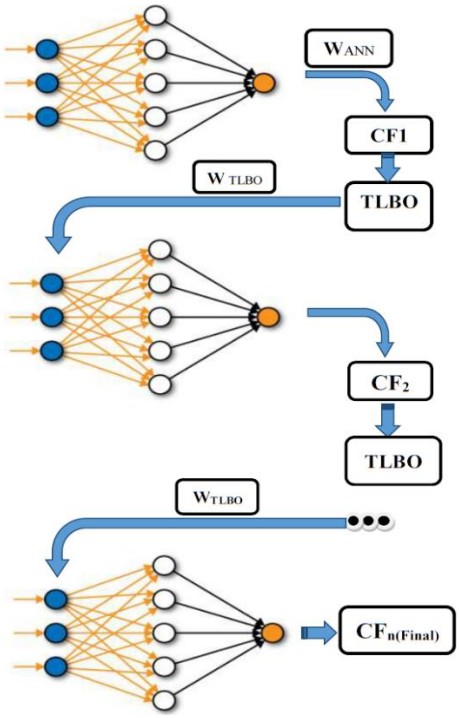
1. feed forward neural network

در این پژوهش تمامی روش‌های مطرح‌شده در بالا، در نرم‌افزار متلب انجام خواهد شد و نتایج با یکدیگر مقایسه می‌شوند.

۲-۴- الگوریتم بهینه‌سازی

در این پژوهش از الگوریتم یادگیری معلم و دانش‌آموز استفاده شده است. در پایان بخش به بیان متغیرها و بخش‌های مختلف استفاده‌شده در این الگوریتم پرداخته شده است. در این الگوریتم، تعداد افراد شرکت‌کننده و یا همان جمعیت اولیه، ۵۰ نفر در نظر گرفته شده که البته قابل تغییر نیز هست. همچنین، تعداد متغیرها برابر (تعداد نورون=n) 2n+n+n+1 بوده که با توجه به ویژگی شبکه‌های پیش‌خور، مقدار 2n برای وزن‌های ورودی است که به تعداد ویژگی ورودی در نورون‌ها باید باشد و نیز هر n بعدی به ترتیب برای بایاس‌های ویژگی‌ها و نیز تعداد خروجی‌ها و عدد یک انتهایی نیز به دلیل داشتن یک خروجی برای یک بایاس خروجی در نظر گرفته شده است. بعد از داشتن ورودی‌ها و متغیرهای الگوریتم TLBO، مقدار بهینه برای ضرایب را یافته و به‌عنوان خروجی به شبکه بازگردانده و مقدار خروجی‌ها محاسبه خواهد شد. در الگوریتم ارائه‌شده، با توجه به اینکه هدف ترکیب شبکه عصبی و الگوریتم TLBO است، باید بتوان این دو الگوریتم را به‌درستی در کنار یکدیگر قرار داد. الگوریتم بهینه‌سازی به‌تنهایی قادر به پیش‌بینی نبوده و الگوریتم شبکه عصبی نیز به‌تنهایی قادر به بهبود نتایج نخواهد بود. در این پژوهش تلاش شده است تا با کنار یکدیگر قرار دادن این دو الگوریتم تا حدودی بتوان این دو ضعف را در الگوریتم‌های بیان‌شده از بین برد و درنهایت به الگوریتم مناسب که قادر به بهینه‌سازی و پیش‌بینی باشد، دست یافت. بدین منظور با توجه به اینکه در شبکه عصبی، یال‌های شبکه اوزان تشخیصی را به‌عنوان خروجی داده و میزان دقت در تشخیص‌ها نیز دقیقاً به این اندازه‌ها مربوط است، تصمیم گرفته شد که این یال‌ها را ابتدا در مرتبه اول شبکه عصبی تولید نموده و در اختیار الگوریتم TLBO قرار دهد و سپس، دیگر شبکه عصبی با این اوزان هیچ کاری نداشته باشد و این الگوریتم بهینه‌سازی باشد که این اوزان را با توجه به تابع هزینه در نظر گرفته‌شده که همان خطای میان تشخیص شبکه عصبی و مقادیر واقعی است، بهبود دهد. هر بار که الگوریتم بهینه‌سازی به مقادیر جدیدی برسد، این اوزان وارد شبکه عصبی شده و فرایند تشخیص صورت می‌گیرد و اختلاف میان واقعیت و تشخیص از طریق تابع هزینه محاسبه خواهد شد؛ تا جایی این کار پیش خواهد رفت که دیگر مقادیر بهتری به دست نیاید.

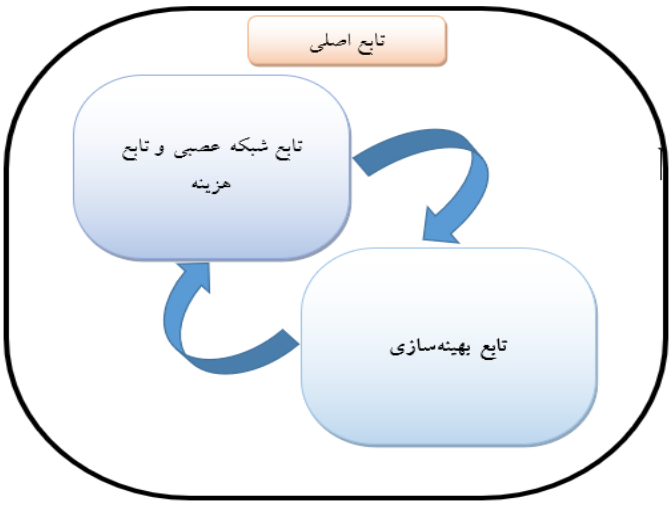
در شکل ۱ می‌توان فرآیند بررسی و بهینه‌سازی در راستای رسیدن به تشخیصی دقیق‌تر را مشاهده نمود. در موارد موجود با یک بار تشخیص شبکه عصبی، امر تشخیص به انتها می‌رسد اما در این پژوهش، ابتدا داده‌ها در اختیار شبکه عصبی قرار گرفته و شبکه عصبی حدس اولیه خود را می‌زند، اوزان در نظر گرفته‌شده برای معادله رگرسیونی تشخیص را ارائه داده و این اوزان وارد الگوریتم TLBO می‌شود. در این الگوریتم، اوزان بر اساس ماهیت و روند یافتن پاسخ بهینه در الگوریتم بهینه‌سازی مذکور، تغییر داده شده و برای بررسی میزان نزدیکی به واقعیت مجدداً وارد شبکه عصبی مصنوعی شده اما این بار دیگر پیش‌بینی صورت نخواهد گرفت و تنها معادله رگرسیونی جدید و میزان نزدیکی جواب به واقعیت مورد بررسی قرار می‌گیرد و پاسخ‌ها برای جهت‌دهی به روند بهینه‌سازی مجدد به الگوریتم بهینه‌سازی وارد می‌شود. این امر به همین ترتیب ادامه خواهد یافت تا خطای تشخیص به تکرار خاصی (شرط خروج) رسیده و میزان دقت به حد مطلوب نزدیک شده باشد.



شکل ۱. نمودار الگوریتم شبکه عصبی

۳-۴- عملکرد الگوریتم

در این الگوریتم، درمجموع سه تابع کدنویسی شده است؛ در تابع اول با نام تابع مادر، عملیات مربوط به شبکه و نیز فراخوانی تابع هزینه و نیز الگوریتم بهینه‌سازی صورت می‌گیرد. در تابع دوم، تابع هزینه موردنیاز برای محاسبه مقدار برازندگی تحلیل به الگوریتم معرفی می‌شود که خود، تابعی از شبکه خواهد بود. در انتها نیز الگوریتم TLBO فراخوانی و مقادیر بهینه برای هر یک از ضرایب محاسبه خواهد شد. در شکل ۲ می‌توان نحوه عملکرد الگوریتم ارائه‌شده را مشاهده نمود.



شکل ۲. نحوه عملکرد الگوریتم ارائه‌شده

بخش اول داده‌ها به طُرقی که در بخش‌های قبل بیان شده است، وارد الگوریتم می‌شود. سپس، مختصات و نوع شبکه عصبی مشخص شده و داده‌های موردنیاز شبکه وارد الگوریتم خواهد شد. برای دور ابتدایی باید میزان درصد

آموزش و تست نیز مشخص گردد اما از دورهای بعد نیاز به این کار نبوده زیرا آموزش از طریق الگوریتم بهینه‌سازی صورت می‌گیرد. محاسبه میزان دقت ورودی از طریق تابع هزینه ارائه‌شده که اختلاف میان میزان حدس زده شده‌ها با میزان واقعی داده‌ها بوده و به‌صورت مربع خطا، محاسبه خطا صورت گرفته شده است. در مرحله بعد، وارد نمودن شبکه با اوزان مشخص‌شده توسط شبکه عصبی به الگوریتم TLBO برای تغییر بهینه‌ی اوزان ارائه‌شده صورت خواهد گرفت. در الگوریتم بهینه‌سازی، داده‌های اولیه به الگوریتم داده شده و سپس، تولید جمعیت اولیه که از ساختار الگوریتم TLBO بوده و همان معلمین و دانش‌آموزان هستند، صورت خواهد گرفت. در مرحله بعد، معلمین و دانش‌آموزان مشخص شده و روند یادگیری بر اساس اصولی که در بخش‌های قبل ارائه شده است، صورت خواهد گرفت. بعد از آن، میزان اوزان به‌دست‌آمده وارد شبکه عصبی شده و میزان تابع هزینه ارائه‌شده محاسبه می‌شود. بعد از محاسبه میزان این تابع هزینه، این مقادیر با مقادیر حاصل‌شده از مراحل قبل مقایسه می‌شود و در صورت رسیدن به شرط همگرایی از چرخه خارج خواهد شد. در غیر این صورت، چرخه الگوریتم TLBO تکرار خواهد شد تا این شرط، مطلوب و قابل‌قبول گردد. روند اتخاذشده برای تحلیل که حضور در شبکه عصبی و گرفتن قدرت یادگیری آن و آموزش به‌صورت الگوریتم معلم-دانش‌آموز است، باعث خواهد شد توسعه مدل شبکه عصبی بهتر صورت پذیرد. درواقع، یک شبکه عصبی بهینه ارائه شده است که برای این کار تنها علت حضور در شبکه عصبی اولیه و آموزش به روش عادی نیز نیاز داشتن به ضرایب اولیه صحیح است. بدین منظور، بعد از تهیه داده‌های اولیه که در این پژوهش به‌صورت متنی این کار صورت می‌پذیرد، داده‌ها به‌صورت یک‌جا وارد شده و تماماً به‌صورت داده‌های تست در نظر گرفته می‌شوند و بهینه‌سازی بر اساس ضرایب به‌دست‌آمده اولیه در شبکه عصبی صورت می‌پذیرد و بعد از هر تغییر تمامی داده‌ها بررسی و میزان دقت محاسبه می‌شود. سرانجام، با توجه به ساختار الگوریتم بهینه‌سازی معلم-دانش‌آموز، تولید جمعیت‌های هدفمند برای اوزان در معادلات رگرسیون شبکه عصبی صورت گرفته شده و میزان دقت در چرخه‌های متوالی تا درصدهای بالا پیش می‌رود.

۵- یافته‌ها

۵-۱- آمار توصیفی

در این بخش به بررسی آمار توصیفی داده‌های جمع‌آوری‌شده پرداخته شده است. همان‌طور که قبلاً نیز مطرح شد، داده‌های مربوط به سال‌های ۱۳۹۷ و ۱۳۹۸ جمع‌آوری گردید. در جدول ۱، تمامی این متغیرها مورد ارزیابی قرار می‌گیرند.

جدول ۱. ارزیابی متغیرهای پژوهش

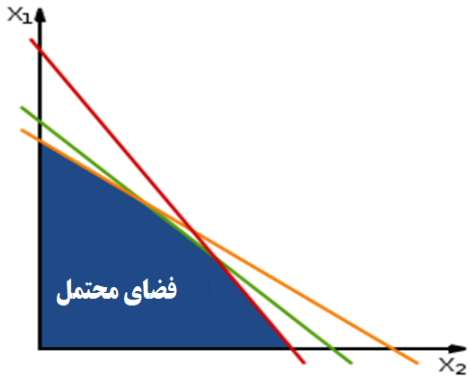
متغیر	سال ۱۳۹۷		سال ۱۳۹۸	
	میانگین	انحراف استاندارد	میانگین	انحراف استاندارد
میزان مبالغ بستانکاری	۳۳۸۰۹۲۳۳۲/۶	۷۵۴۱۴۰۰۳۹/۹	۲۹۵۱۳۶۵۸۹/۸	۵۹۰۵۱۵۳۲۵/۷
میزان مبالغ بدهکاری	۳۳۰۲۸۴۶۰۱	۷۴۶۲۹۵۴۲۱/۵	۳۰۳۳۱۹۲۵۱/۶	۶۰۱۸۴۲۹۹۲/۹
تعداد روزهای تراکنش	۸۷/۴۱۹۳۵۴۸۴	۸۳/۲۶۹۲۸۵۶۴	۸۰/۷۸۴۴۸۲۷۶	۷۹/۴۲۳۲۲۵۸۳
تعداد تراکنش	۲۳۴/۷۷۴۱۹۳۵	۳۰۴/۵۸۴۲۹۱۹	۲۳۷/۴۲۲۴۱۳۸	۳۲۷/۰۷۲۳۴۰۱
مجموع تراکنش	۶۶۸۳۷۶۹۳۳/۵	۱۴۹۹۴۳۱۳۷۹	۵۹۸۴۵۵۸۴۱/۴	۱۱۹۱۰۰۵۰۸۸
تعداد تراکنش‌های بدهکاری	۲۰۵/۷۵۸۰۶۴۵	۲۷۰/۱۴۷۸۷۷۹	۲۰۶/۹۵۶۸۹۶۶	۲۸۸/۴۷۰۳۱۰۲
تعداد تراکنش‌های بستانکاری	۲۹/۰۱۶۱۲۹۰۳	۴۰/۲۷۵۰۶۷۲۱	۳۰/۴۶۵۵۱۷۲۴	۴۹/۹۷۹۷۲۳۰۲

بعد از ارزیابی متغیرهای پژوهش، به بررسی نتایج حاصل از پیاده‌سازی الگوریتم‌های انجام شده پرداخته شده است و نتایج به‌دست‌آمده با یکدیگر مقایسه می‌شوند.

۲-۵- شبکه عصبی ارائه‌شده

همان‌طور که بیان شد، با قرار دادن شبکه عصبی در دل یک الگوریتم بهینه‌سازی می‌توان میزان دقت بهینه‌سازی و اختلاف خطای پاسخ‌ها نسبت به پاسخ واقعی را به‌عنوان تابع هزینه در نظر گرفت و از این طریق، میزان همگرایی پاسخ‌ها به مقدار واقعی را بیشتر نمود. بدین منظور، بر اساس الگوریتم ارائه‌شده تصمیم بر این شد که ضرایب روی یال‌های شبکه عصبی به‌عنوان ضرایب مورد بهینه‌سازی به الگوریتم وارد شده و با تغییرات این ضرایب به سمت مقادیر بهینه سوق داده شود. درواقع، با این کار قدرت یادگیری از شبکه عصبی سلب شده و با روش این پژوهش که همان یادگیری انسان است، به شبکه آموزش داده می‌شود. درنهایت، این ضرایب وارد شبکه عصبی شده و میزان بهبود از طریق تابع برازندگی در نظر گرفته‌شده سنجیده می‌شود.

نکته مهم در مورد بهینه‌سازی این است که در هنگام یادگیری از طریق عادی در یک فضای بی‌نهایت از طریق ورودی‌ها به سمت ضرایب حرکت صورت می‌گیرد اما در بهینه‌سازی با توجه به شرایط و قیود اعمال‌شده بر مسئله، این جستجو در محدوده خاصی صورت می‌گیرد و با رشد و پیشروی در بهینه‌سازی، این محدوده به سمت پاسخ بهینه حرکت کرده و مدام کوچک‌تر می‌شود. در بهینه‌سازی به این فضای ایجادشده و میان قیود بهینه‌سازی، فضای محتمل می‌گویند. برای درک بهتر از این مسئله، در شکل ۳ تصویری از یک فضای محتمل در مسائل ساده برای حل معادلات به روش بهینه‌سازی ارائه شده است.



شکل ۳. نمایش فضای محتمل

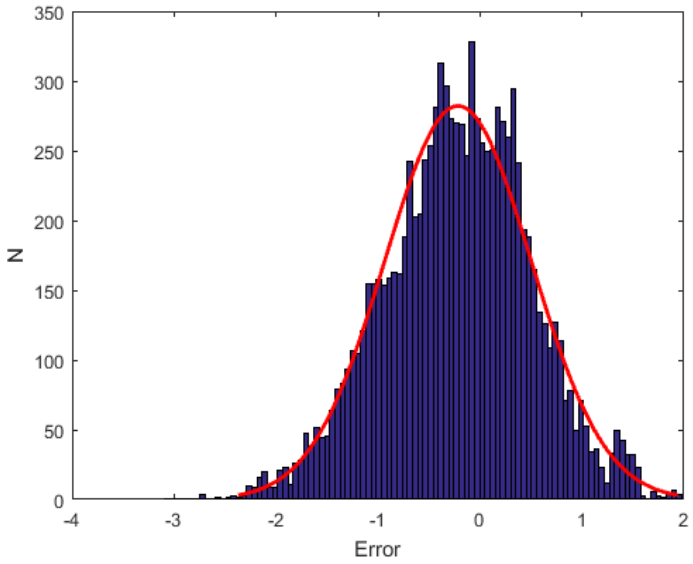
در شکل ۳ مشاهده می‌شود که محیط آبی‌رنگ فضای بین محور افق و عمودی و معادلات خط ترسیم‌شده در صفحه است. این معادلات، قیود مسئله بوده و پاسخ بهینه در این ناحیه قرار دارد و الگوریتم بهینه‌سازی خارج از این ناحیه برای رسیدن به پاسخ بهینه تلاشی صورت نخواهد داد. در تمام مسائل بهینه‌سازی، این ناحیه موجود است اما در مسائل پیچیده نظیر الگوریتم‌های فراابتکاری، این ناحیه به‌سختی قابل تشخیص و ترسیم خواهد بود زیرا قیود حاکم بر مسئله به‌صورت معادلات مشخص نبوده و در هر مرحله این قیود به‌صورت پویا در حال تغییر هستند.

بعد از آشنایی با این ناحیه می‌توان وارد الگوریتم ارائه‌شده در این پژوهش شد اما ابتدا باید پارامترهای بهینه‌سازی مورد استفاده در این الگوریتم ارائه گردد که در جدول ۲ مشاهده می‌شوند.

جدول ۲. مشخصات ورودی الگوریتم

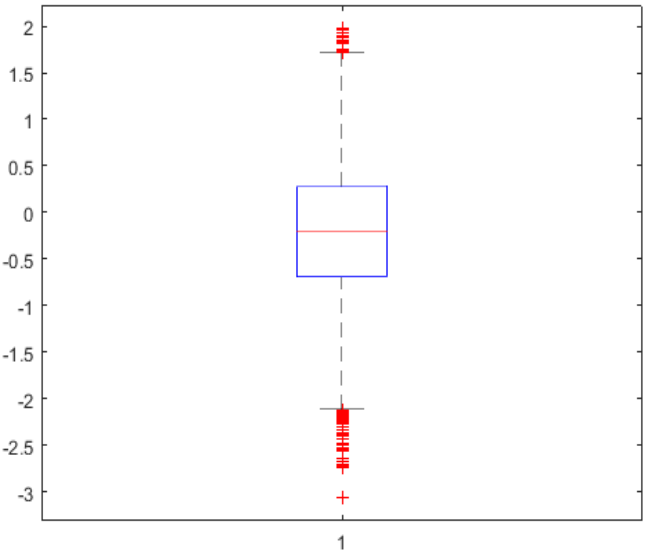
متغیر	مقدار
نام الگوریتم بهینه‌سازی	*TLBO*
نوع شبکه عصبی	شبکه پیش‌خور
تعداد نورون‌ها	۱۰
تابع هزینه	$NMSE = \frac{\overline{(T - Network(I))}}{Variance(T')}$
تابع آموزش شبکه	لونبرگ مارکوتز
تابع عملکرد شبکه	*MSE*
تعداد چرخه بهینه‌سازی	رسیدن به همگرایی در مقدار تابع هزینه
تعداد افراد (متغیر جمعیت اولیه) واردشده به بهینه‌سازی	۵۰

با شروع الگوریتم، در ابتدا ورودی‌ها به شبکه وارد شده و شبکه یک بار آموزش به سبک خود را صورت داده و اوزان را ارائه می‌دهد. در مرحله بعد، بر اساس مقادیر حاصله، میزان تابع هزینه در این الگوریتم محاسبه شده و به‌عنوان خطای ورودی به الگوریتم معرفی می‌شود. در ادامه می‌توان میزان خطا را مشاهده نمود.



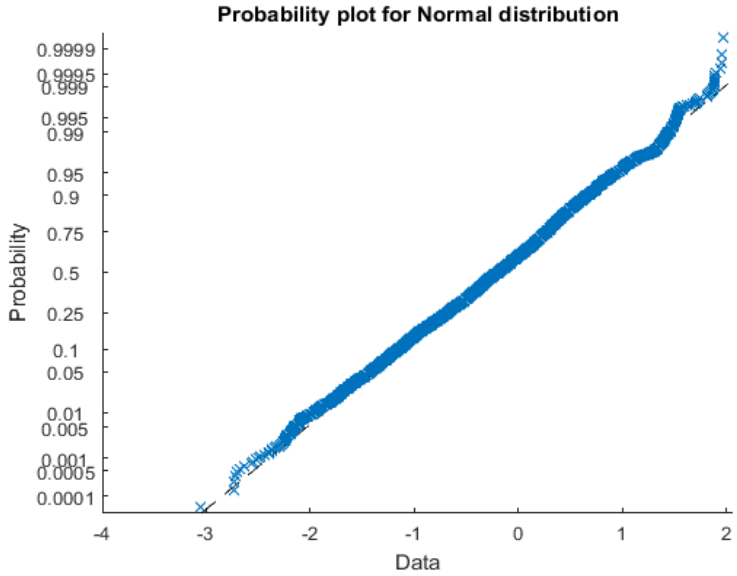
شکل ۴. نمودار هیستوگرام توزیع خطا در شبکه عصبی ورودی

در نمودار شکل ۴ مشاهده می‌شود که تعداد فراوانی، در بازه بزرگی از خطاها غیر صفر بوده و نمودار دارای قله غیر واحد بوده است که این نشان از خطای زیاد در تشخیص دارد. در صورت مناسب بودن، بهینه‌سازی بین این قله تیزتر شده و از میزان چولگی این نمودار کاسته می‌شود و برعکس، به میزان کشیدگی این نمودار افزوده می‌گردد. مقادیر کشیدگی و چولگی در نمودار خطای ورودی برابر ۳/۰۵۳۹ و ۰/۱۱۵۹- به دست آمده است. همچنین برای مشخص شدن میزان تمرکز خطاها و میزان خطاهای پرت می‌توان از نمودار جعبه‌ای استفاده نمود که برای داده‌های ورودی در این پژوهش به‌صورت شکل ۵ است. داده‌های پرت در این نمودار خارج از بازه جعبه قرار گرفته و هرچه میزان این اعداد بیشتر شود، نشان از خطای محاسبات است.



شکل ۵. چارک‌های اول، دوم و سوم خروجی تحلیل و داده‌های پرت

در شکل ۵ مشاهده می‌شود که میزان بازه خطا در بازه ۱/۵ الی ۲ بوده و داده‌هایی که با علامت قرمز مشخص شده‌اند، همان خطاهای خارج از بازه مناسب برای این داده‌ها هستند. همچنین، میانه در این خطاها کمتر از صفر و در حدود ۰/۵ شده است. این امر نشان از این دارد که تمرکز داده‌های خطا هنوز در گرایش به صفر موفق نبوده است و مرکزیت داده‌ها کمتر از صفر حاصل شده است. در ادامه، بررسی میزان دقت تشخیص در شبکه را از طریق نمودار توزیع احتمالاتی مورد بررسی قرار داده و میزان دقت و پراکندگی را مشاهده خواهیم نمود.

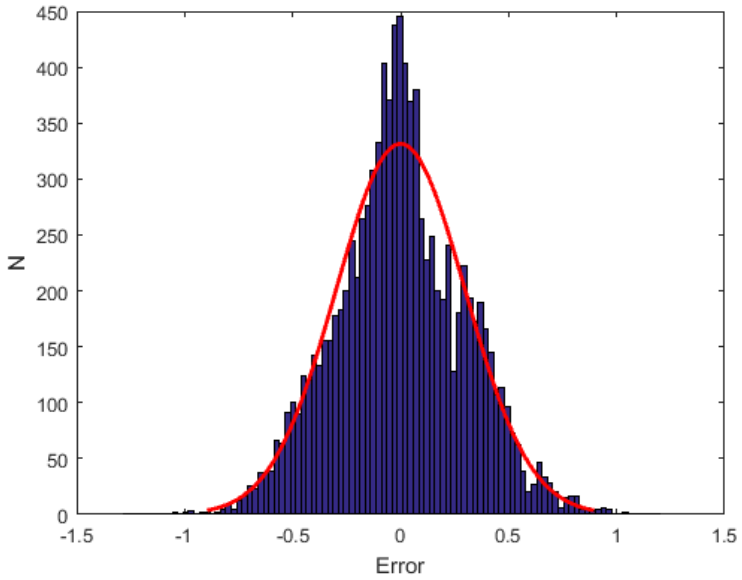


شکل ۶. نمودار توزیع احتمالاتی خطا

در شکل ۶ مشاهده می‌شود که در نواحی مرکزی، نمودار نقاط بر روی خط تعادل و بازشده قرار دارد ولی در نواحی بیرونی، نمودار پراکندگی و میزان انحراف معیار بیشتر شده است و این مقادیر از میزان استاندارد فاصله گرفته‌اند. اگر در این زمینه به نمودار جعبه‌ای مراجعه گردد، مشاهده خواهد شد که در این نواحی، داده‌های پرت قرار داشته و

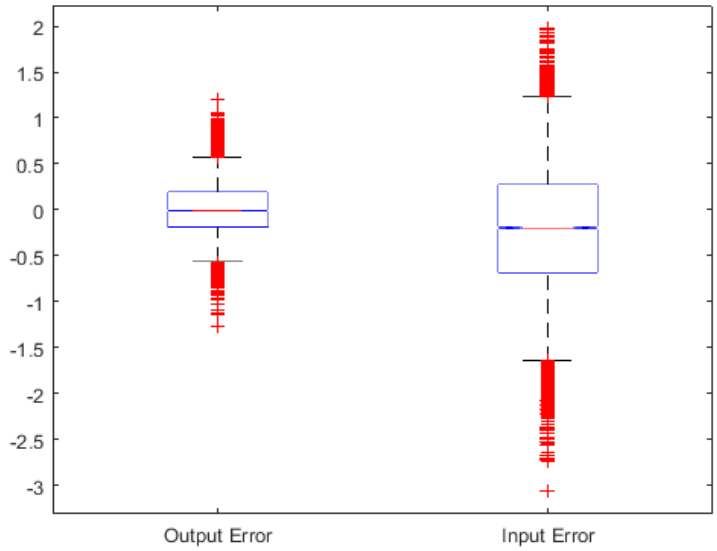
به همین دلیل است که در نمودار توزیع احتمالاتی نیز در این نواحی پراکندگی مشاهده شده و به همین دلیل از میزان دقت در این مرحله کاسته شده است.

در ادامه به بررسی پاسخ نهایی الگوریتم ارائه‌شده پرداخته شده است. در مرحله اول به بررسی نمودار تغییرات فراوانی و یا همان نمودار هیستوگرام خطا پرداخته شده است. در قیاس با نمودار مرتبط با ورودی الگوریتم، مشاهده می‌شود که به‌وضوح دامنه تغییرات به‌شدت بهبود یافته و تمرکز بیشتر بر روی میزان مبدأ مختصات بوده است. قله این نمودار نوک‌تیز بوده و دیگر خطا در طول داده‌ها توزیع نشده است و اغلب خطای بسیار کمی در قیاس با مقدار مرجع از خود نشان داده‌اند.



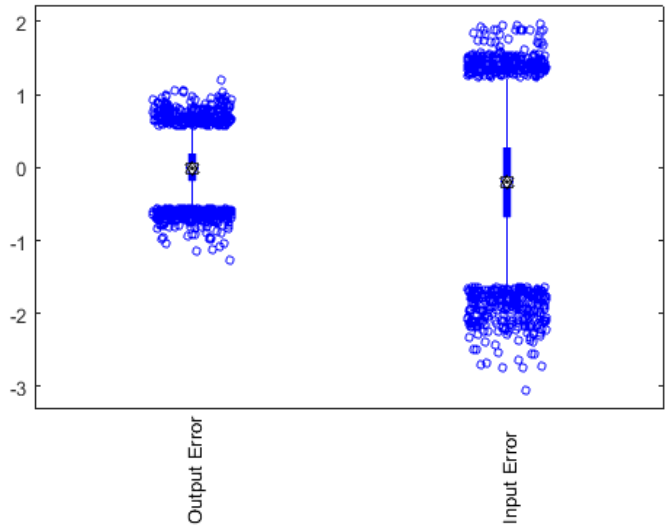
شکل ۷. نمودار هیستوگرام توزیع خطا برای شبکه عصبی خروجی

در این نمودار نیز با بررسی میزان کشیدگی و چولگی می‌توان بررسی مناسب‌تری را برای تغییرات خطا ارائه نمود. مقادیر کشیدگی و چولگی در این مدل، به ترتیب برابر ۳/۱۱۹۷ و ۰/۰۷۰۸ حاصل شده که نسبت به مدل ورودی کمتر شده است. میزان کشیدگی نشان از این است که تمرکز بر روی قله بیشتر شده و میزان کشیدگی در نمودار بالا رفته است. میزان چولگی نیز نشان می‌دهد که نمودار جمع شده و بازه خطا، کوچک‌تر شده است. نکته مهم این است که اختلاف بین این فاکتورها در ورودی و خروجی الگوریتم، اختلاف خطا را نشان نمی‌دهد و به‌صورت نسبی است. در ادامه برای بررسی میزان دامنه تغییرات و تمرکز و میانه خطاها از نمودار جعبه‌ای برای این اعداد استفاده شده است.



شکل ۸. مقایسه خروجی‌های اول، دوم و سوم و داده‌های پرت شبکه عصبی ورودی و خروجی

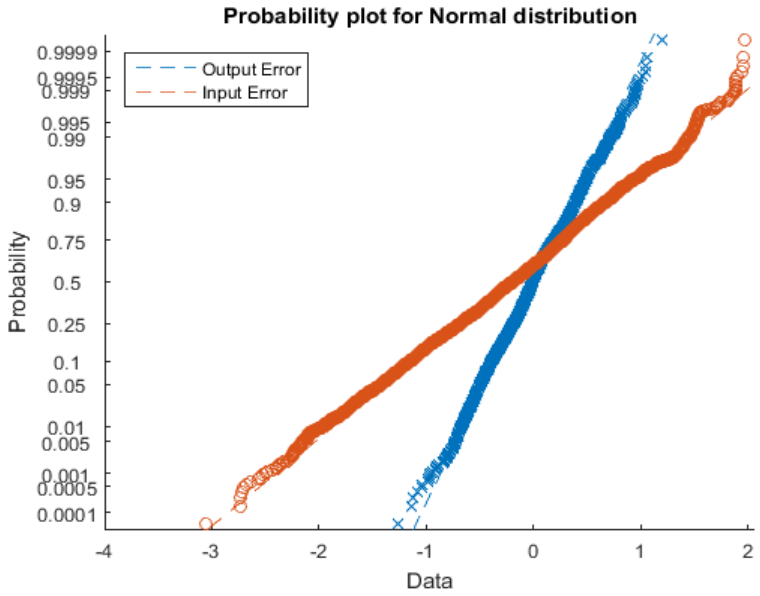
در شکل ۸ مشاهده می‌شود که میزان دامنه تغییرات میان چارک اول و سوم در نمودارهای جعبه‌ای خطا در خروجی به‌شدت کاهش داشته و همچنین، میانه در خطای خروجی در نزدیک به صفر و در خطای ورودی غیر صفر است. این نکته نیز بیانگر بهبود در بهینه‌سازی خواهد بود. داده‌های پرت در هر دو گروه مشاهده شده است اما داده‌های پرت عامل دقیقی در تحلیل نیست زیرا با افزایش دقت در تشخیص داده پرت نیز سخت‌گیری بیشتری اعمال خواهد شد. به همین دلیل تنها می‌توان به حالات دوری داده‌ها اشاره نمود که داده پرت با فاصله زیاد در نمودار خطای خروجی مشاهده نشده است. با وجود این، برای بررسی دقیق‌تر میزان پراکندگی داده‌های پرت نیز تحلیل صورت گرفته شده است. در نمودار شکل ۹ می‌توان به‌وضوح نحوه توزیع و حجم داده‌های پرت را مشاهده نمود.



شکل ۹. مقایسه داده‌های پرت شبکه عصبی ورودی و شبکه عصبی خروجی

در نمودار شکل ۹ مشاهده می‌شود که در هر دو نمودار داده‌های پرت موجود است. در این نمودارها مشاهده شده است که میزان داده‌های دورتر داده‌های خطای ورودی بیشتر بوده است و داده‌های پرت در خطای خروجی بیشتر فشرده بوده و نزدیک به کرانه پایین نمودار جعبه‌ای قرار دارد که بیانگر بهبود در توزیع نیز بوده است. البته، توزیع

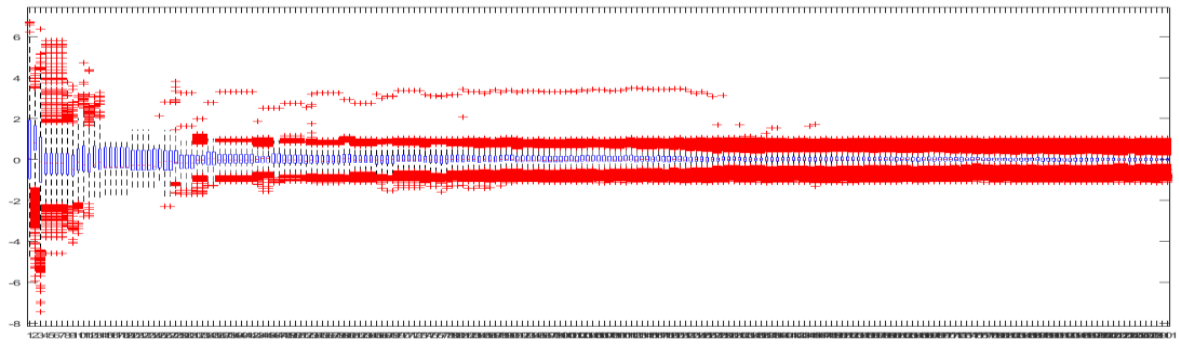
داده‌های خطا دلیلی بر بهبود یا عدم بهبود نیست و تنها در این بهینه‌سازی، این مورد نیز مشاهده شده است. در ادامه، بررسی تغییرات توزیع احتمالاتی خطا در خروجی پرداخته شده و برای درک بهتر، دو نمودار خروجی در یک نمودار ترسیم شده است تا تغییرات به‌صورت مشخص‌تر درک گردد.



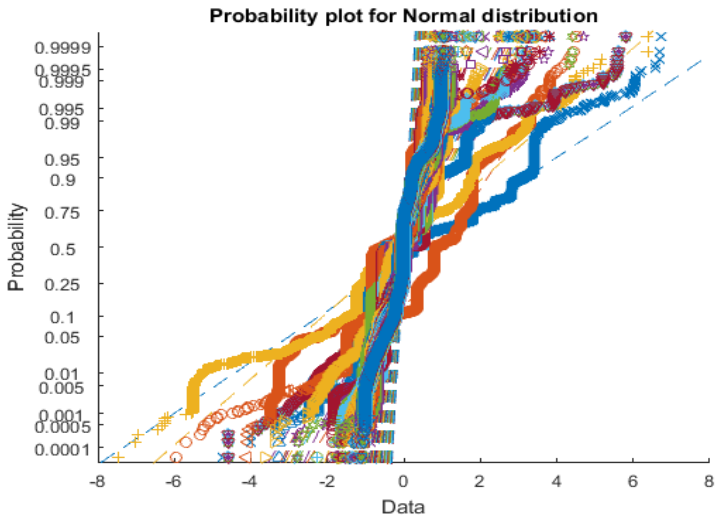
شکل ۱۰. مقایسه توزیع احتمالاتی خطا دو شبکه عصبی

در نمودار شکل ۱۰ مشاهده می‌شود که به‌وضوح میزان دقت در تشخیص در نمودار خطای خروجی قابل تشخیص است. دلیل این نکته این است که میان خروجی خطای داده‌ها بر روی خط پردازش داده‌ها در نمودار خطای خروجی بیشتر شده است. دلیل دیگر این است که میزان شیب در نمودار خروجی در قیاس با ورودی به‌شدت تغییر داشته و شیب در خروجی بیشتر شده است. ازآنجاکه شیب قائم نشان از خطای صفر خواهد بود، می‌توان بیان نمود که خروجی مقداری بسیار مناسب بوده است. دلیل اینکه شیب قائم بهترین پاسخ خواهد بود، این است که با دقت در نمودار فوق مشاهده می‌شود که طول نمودار برای هر ورودی، همان بازه تغییرات خواهد بود و درصورتی‌که بازه تغییرات صفر شود، نشان از این خواهد بود که تمام اعداد برابر و مساوی صفر شده است. به همین دلیل، شیب قائم بر روی نقطه صفر بهترین مقدار خواهد بود. در ادامه، بررسی کلی تغییرات در طول بهینه‌سازی در دستور کار قرار گرفته شده و تلاش شده است تا تغییرات در طول بهینه‌سازی به‌صورت مناسبی نشان داده شود.

برای این منظور، در ابتدا تغییرات توزیع پاسخ بر اساس نمودارهای جعبه‌ای بررسی شده و این نمودارهای تمام چرخه‌ها در کنار یکدیگر ترسیم شده‌اند و می‌توان در طی آن مشاهده نمود که تغییرات میانه و روند میل به خطای صفر چگونه رخ داده شده است. در نمودار شکل ۱۱، می‌توان تغییرات خطا در طول بهینه‌سازی را مشاهده نمود. در این تغییرات می‌توان مشاهده نمود که در حین بهینه‌سازی ناحیه مرکزی که همان باکس مرکزی خطا است، به‌مرور کاهش یافته و در طول بهینه‌سازی با گذر زمان در برخی از مراحل میزان ورودی داده‌های پرت در بهینه‌سازی زیاد شده است ولی همان‌طور که بیان شد، میزان داده پرت از اهمیت چندانی برخوردار نبوده ولی زیاد شدن این داده‌ها باعث کم شدن دقت نیز خواهد شد. درنهایت، مشاهده شده است که بهینه‌سازی ارائه‌شده باعث شده است تا حتی این داده‌های پرت نیز به کمترین میزان خود برسد زیرا این داده‌های پرت بر روی میزان خطا تأثیر اندکی خواهد گذاشت که در ربع چهارم بهینه‌سازی به بهینه‌سازی بر روی داده‌های پرت پرداخته شده است و به‌وضوح قابل مشاهده است که داده‌های پرت در این ناحیه از بین رفته‌اند.

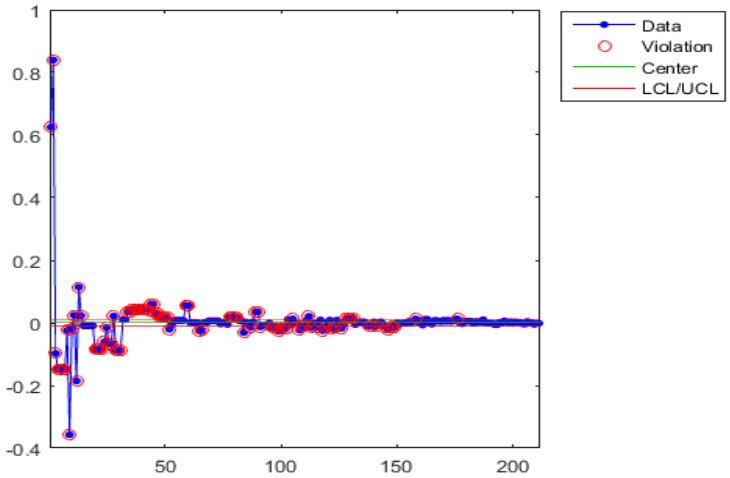


شکل ۱۱. روند تغییرات دقت تشخیص شبکه در طول بهینه‌سازی



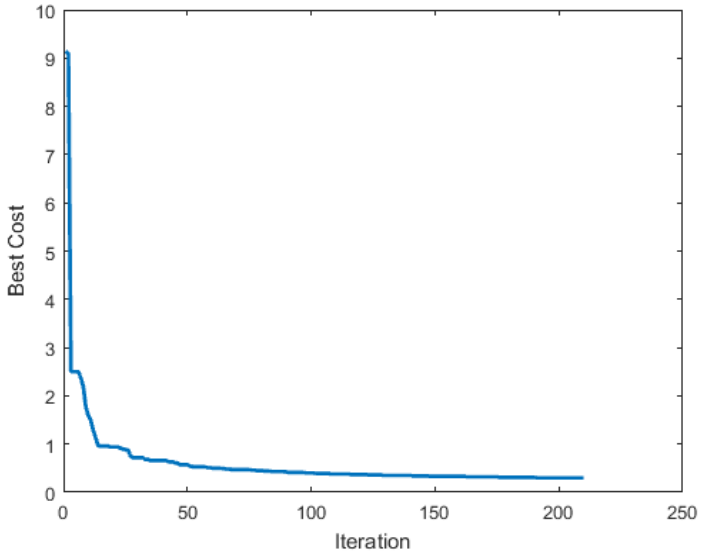
شکل ۱۲. تغییرات توزیع احتمالاتی دقت شبکه در چرخه‌های بهینه‌سازی

در نمودار شکل ۱۲ مشاهده می‌شود که به‌وضوح شیب به حالت عمودی درآمده و بازه خطا در طول بهینه‌سازی کاهش یافته و درنهایت نیز در نزدیکی خط عمود این بهینه‌سازی توقف داشته است. برای بررسی تغییرات خطا در میان تمام داده‌ها، در هر دور یک برآیند مناسب گرفته شده و در مقابل بین تمام برآیندهای همگراشده مقایسه صورت می‌گیرد. سرانجام می‌توان به نمودار ارائه‌شده در شکل ۱۳ رسید که نشان می‌دهد الگوریتم ارائه‌شده در کدام مرحله به بازه مناسب برای پاسخ نهایی رسیده است.



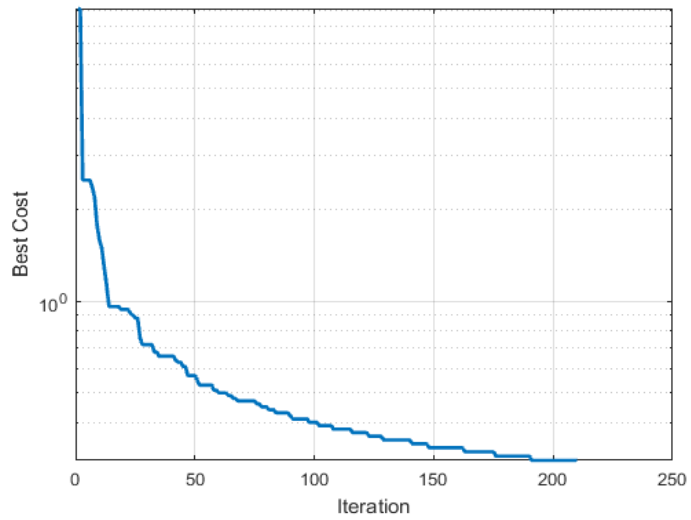
شکل ۱۳. میزان تجاوز خطای تشخیص در بهینه‌سازی

نقاط قرمز که خارج از بازه تعیین‌شده قرار داشته باشند، به‌عنوان پاسخ غیرقابل‌قبول از منظَر همگرایی مناسب در نظر گرفته خواهد شد. مشاهده می‌شود که در انتهای بهینه‌سازی، این خروج از بازه تعیین‌شده به نسبت بسیار کمی نسبت به ابتدای بهینه‌سازی اتفاق افتاده است. نمودار تغییرات تابع هزینه نیز در طول بهینه‌سازی ترسیم شده است که در شکل ۱۴ می‌توان این نمودار را مشاهده نمود.



شکل ۱۴. روند بهبود تشخیص در طول بهینه‌سازی

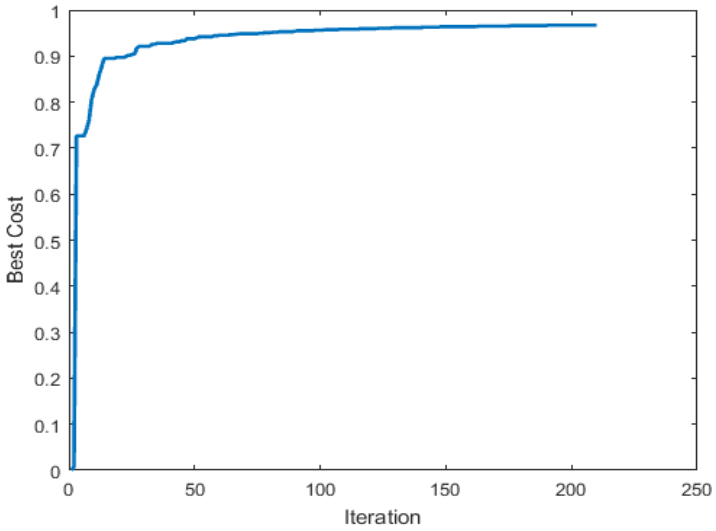
در نمودار شکل ۱۴ مشاهده می‌شود که مقادیر خطا در طول بهینه‌سازی به‌مرور کاهش یافته است. این تغییرات به مرور زمان کاهش یافته است. دلیل این نحوه تغییرات نیز این است که فضای محتمل در ابتدا بزرگ بوده و رشد بهینه‌سازی با سرعت صورت خواهد گرفت اما با گذر بهینه‌سازی و افزایش دقت که از طریق میل به پاسخ بهینه است، این میل که ناشی از تغییرات فضای محتمل است، باعث خواهد شد تا این رشد کاهش یابد زیرا مدام این فضا کوچک‌تر شده و یافتن پاسخ بهتر با سختی بیشتری صورت می‌گیرد. در ادامه برای درک بهتر تغییرات در طول بهینه‌سازی، نمودار لگاریتمی تغییرات خطا در طول بهینه‌سازی ارائه شده است.



شکل ۱۵. روند لگاریتمی بهبود عدم تشخیص در طول بهینه‌سازی

در نمودار لگاریتمی ارائه‌شده در شکل ۱۵ مشاهده شده است که تغییرات به‌صورت مناسبی قابل مشاهده است ولی با وجود لگارتیمی شدن مقادیر خطا، هنوز در ابتدای الگوریتم شیب زیاد و به‌مرور این مقادیر با شیب کمتری کاهش یافته است.

در ادامه، بررسی تغییرات راندمان بهینه‌سازی در مقایسه با ورودی الگوریتم ارائه شده است که می‌توان این تغییرات را به‌صورت نموداری مشاهده نمود.



شکل ۱۶. روند بهبود تشخیص در طول بهینه‌سازی

در شکل ۱۶ مشاهده می‌شود که راندمان الگوریتم به حد بسیار مطلوبی رسیده است و نسبت به میزان ورودی که شبکه عصبی وارد بهینه‌سازی نموده، بهینه‌سازی حدود ۹۰ درصدی را صورت داده است که درصد بسیار مطلوب و مناسبی برای این حجم از داده محسوب می‌شود.

۶- بحث و نتیجه‌گیری

پیش‌بینی و تشخیص مشتریان وفادار از غیر وفادار، دغدغه و مسئله اصلی هر مؤسسه مالی به شمار می‌رود. شناسایی مشتریان وفادار به این سازمان‌ها و مؤسسات، می‌تواند باعث بهبود عملکرد و پیشرفت آن‌ها گردد. ازاین‌رو، شناسایی این افراد از اولویت بالایی برای این مؤسسات برخوردار است. داده‌های مورد استفاده در این پژوهش، از یک بانک داخلی جمع‌آوری شده است که شامل فیلدهای میزان مبالغ بستانکاری، میزان مبالغ بدهکاری، تعداد روزهای تراکنش، تعداد تراکنش، مجموع مبالغ تراکنش، تعداد تراکنش‌های بدهکاری و تعداد تراکنش‌های بستانکاری در سال‌های ۱۳۹۷ و ۱۳۹۸ است. بعد از بهبود عملکرد شبکه عصبی، نتایج زیر حاصل گردید:

مهم‌ترین نتیجه حاصل‌شده از این پژوهش این است که میزان دقت تشخیص در این پژوهش بسیار مناسب و در حدود ۹۷ درصد شده است. البته، این دقت در مقایسه با نتایج حاصل‌شده از تحقیق ژو و همکاران[1] (۲۰۱۹) که میزان دقت تشخیص در آن برابر با ۹۴/۳ درصد بود، افزایش داشته است.

نتایج نشان داد که با گذر زمان از بهینه‌سازی میزان داده‌های پرت در تشخیص نیز بسیار کاسته شده است. درواقع، داده‌های پرت برخی داده‌های خاص در دیتاست بوده که از الگوریتم و روند خاصی پیروی نمی‌کنند. الگوریتم بعد از بهینه‌سازی از طریق ارائه رویکرد مناسب توانست این مقادیر را تا حدود زیادی تشخیص دهد.

1. Zhou et al.

در الگوریتم ارائه داده‌شده نشان داده شد که بهینه‌سازی در تمام موارد همگرایی تا ۹۹ درصد بوده است و جدا از ورودی‌هایی که شبکه عصبی به ورودی بهینه‌سازی ارائه می‌نماید، عمل نموده است زیرا در شبکه عصبی با توجه به انتخاب تصادفی داده‌های آموزش و آزمودن ممکن است دقت در تشخیص‌ها تغییر داشته باشد که این تغییرات بر روی عملکرد الگوریتم تأثیر چندانی را نگذاشته است.

در نمودارهای توزیع نشان داده شد که هرچه کشیدگی در نمودار، کم و نقطه عطف بر روی نقطه صفر باشد، عملکرد بهتری ارائه می‌شود. درواقع، باید به‌صورت حرف S جمع باشد که عطف بر صفر بیانگر مرکزیت داده‌های بر روی خطای صفر و جمع بودن این نمودار نشان از کم بودن بازه تغییرات خطا است.

در نمودار هیستوگرام نیز نشان داده شد که برخلاف بررسی توزیع داده‌ها که کشیدگی و چولگی کم، معیار مناسبی نخواهد بود، در این بررسی با توجه به اینکه نمودار هیستوگرام خطا ترسیم می‌شود، کشیدگی بیشتر بر روی نقطه صفر نشان از این است که میزان خطای صفر تکرار زیادی داشته و هرچه چولگی کم شود، خطای توزیع‌شده در سطح بهینه‌سازی کم شده است.

منابع

Abdar, M., Zomorodi-Moghadam, M., & Zhou, X. (2018). An ensemble-based decision tree approach for educational data mining. In *2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC)* (pp. 126-129). IEEE.

Bahari, T. F., & Elayidom, M. S. (2015). An efficient CRM-data mining framework for the prediction of customer behaviour. *Procedia computer science*, *46*, 725-731.

Chen, H., Fuller, S. S., Friedman, C., & Hersh, W. (Eds.). (2006). *Medical informatics: knowledge management and data mining in biomedicine* (Vol. 8). Springer Science & Business Media.

Das, S., & Nayak, J. (2022). Customer segmentation via data mining techniques: state-of-the-art review. *Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021*, 489-507.

Dash, G. P., & Nayak, B. (2022). The Impact of Customer Relationship Management (CRM) in the Banking Sector Using Machine Learning.

Godin, M. (2002). Health care informatics: An interdisciplinary approach. *CIN: Computers, Informatics, Nursing*, *20*(6), 217-219.

KA, S., & Johnson, B. (2017). Electronic customer relationship management (e-CRM) techniques for effective banking. *Finance India*, *31*(1).

Kumar, P., & Mokha, A. K. (2022). Electronic customer relationship management (E-CRM) and customer loyalty: The mediating role of customer satisfaction in the banking industry. *International Journal of E-Business Research (IJEBR)*, *18*(1), 1-22.

Latheef, J., & Vineetha, S. (2021a). Exploring Data Visualization to Analyze and Predict Customer Loyalty in Banking Sector with Ensemble Learning. *International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)*, *4*(9).

Latheef, J., & Vineetha, S. (2021b). Predicting customer loyalty in banking sector with mixed ensemble model and hybrid model. In *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics, Volume 2* (pp. 363-371). Springer Singapore.

LaTour, K. M., & Eichenwald, S. (2010). *Health information management: Concepts, principles, and practice* (pp. 544-545). P. K. Oachs (Ed.). Ahima.

Manyanga, W., Viriri, P., Mashonjowa, P., Ruzvidzo, P., & Manyanga, T. (2022). The impact of electronic customer relationship management on customer retention in zimbabwe's banking sector. *International Journal of Information, Business and Management*, *14*(3), 55-66.

Mining, W. I. D. (2006). Data mining: Concepts and techniques. *Morgan Kaufinann*, *10*(559-569), 4.

Obenshain, M. K. (2004). Application of data mining techniques to healthcare data. *Infection Control & Hospital Epidemiology*, *25*(8), 690-695.

Rezaei, M., Sanayei, A., Amiri Aghdaie, S. F., & Ansari, A. (2022). Improving the Omnichannel customers' lifetime value using association rules data mining: a case study of agriculture bank of Iran. *Interdisciplinary Journal of Management Studies (Formerly known as Iranian Journal of Management Studies)*, *15*(1), 49-68.

Tan, J. (2008). *Medical informatics: Concepts, methodologies, tools, and applications*. Hershey: IGI Global snippet.

Tao, X., Zhou, X., Zhang, J., & Yong, J. (2016). Sentiment analysis for depression detection on social networks. In *Advanced Data Mining and Applications: 12th International Conference, ADMA 2016, Gold Coast, QLD, Australia, December 12-15, 2016, Proceedings 12* (pp. 807-810). Springer International Publishing.

Wager, K. A., Lee, F. W., & Glaser, J. P. (2005). *Managing health care information systems: a practical approach for health care executives*. John Wiley & Sons.

Zhou, X., Bargshady, G., Abdar, M., Tao, X., Gururajan, R., & Chan, K. C. (2019). A case study of predicting banking customers behaviour by using data mining. In *2019 6th international conference on behavioral, economic and socio-cultural computing (BESC)* (pp. 1-6). IEEE.

Zhou, X., Tao, X., Rahman, M. M., & Zhang, J. (2017). Coupling topic modelling in opinion mining for social media analysis. In *Proceedings of the international conference on web intelligence* (pp. 533-540).

Zhou, X., Wang, Y., Tsafnat, G., Coiera, E., Bourgeois, F. T., & Dunn, A. G. (2015). Citations alone were enough to predict favorable conclusions in reviews of neuraminidase inhibitors. *Journal of clinical epidemiology*, *68*(1), 87-93.

استناد به این مقاله: نامنی، سمیرامیس. (۱۴۰۳). استخراج ویژگی‌های مناسب جهت تحلیل رفتار و پیش‌بینی دلایل ریزش مشتریان بانک‌ها. *فصلنامه پژوهش‌های نوین در شهر هوشمند*، ۳(۱)، ۲۹-۴۹.

New Researches in The Smart City is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.